



Pragmatism and the predictive mind

Daniel Williams¹ 

© The Author(s) 2018. This article is an open access publication

Abstract Predictive processing and its apparent commitment to explaining cognition in terms of Bayesian inference over hierarchical generative models seems to flatly contradict the pragmatist conception of mind and experience. Against this, I argue that this appearance results from philosophical overlays at odd with the science itself, and that the two frameworks are in fact well-poised for mutually beneficial theoretical exchange. Specifically, I argue: *first*, that predictive processing illuminates pragmatism’s commitment to both the primacy of pragmatic coping in accounts of the mind and the profound organism-relativity of experience; *second*, that this pragmatic, “narcissistic” character of prediction error minimization undermines its ability to explain the distinctive *normativity* of intentionality; and *third*, that predictive processing therefore *mandates* an extra-neural account of intentional content of exactly the sort that pragmatism’s communitarian vision of human thought can provide.

Keywords Predictive processing · The free energy principle · Predictive coding · Pragmatism · Intentionality · Organism-relativity · Mental representation · Communitarian · Bayesian brain

1 Introduction

Brains are prediction machines that restlessly seek to match incoming sensory inputs against internally generated, model-based predictions of such inputs. Or so says “predictive processing,” a recent overarching theory of neural function in cognitive and computational neuroscience (Clark 2013, 2016; Friston 2010; Hohwy 2013). In its most ambitious formulation, this neural function is itself a special case of a more fundamental imperative in biological systems to maintain themselves within the narrow subset of biophysical states required for homeostasis (Friston 2009, 2010). Given its

✉ Daniel Williams
dw473@cam.ac.uk

¹ Faculty of Philosophy, Trinity Hall, University of Cambridge, Cambridge, UK

explanatory scope and far-reaching consequences for our understanding of the nature of life and mind, predictive processing is currently enjoying an enormous amount of attention in both the scientific and philosophical literature.

As some have noted (e.g. Hohwy 2013), the emphasis on (approximate) Bayesian inference over hierarchically structured generative models at the heart of this emerging framework appears to straightforwardly contradict the pragmatist conception of mind and experience. In this paper, I argue that this apparent tension rests on philosophical overlays not motivated by the science itself, and that the two frameworks are in fact ideally positioned for mutually beneficial theoretical exchange.

I structure the paper as follows.

In Section 2, I identify three fundamental and unifying commitments in the pragmatist conception of mind and experience: the primacy of pragmatic coping in accounts of the mind, the organism-relativity of experience, and the social construction of human thought. In Section 3, I articulate the broadly “Cartesian” presentation of predictive processing as advanced by Hohwy (2013; 2014; Kiefer and Hohwy 2017) and explain how it contradicts each of these commitments. In the rest of the paper, I then argue that this Cartesian interpretation mischaracterises predictive processing, and I explain how the two frameworks can in fact positively illuminate one another.

In Section 4, I argue that predictive processing both *vindicates* and *illuminates* the first two pragmatist commitments outlined in Section 2: namely, the primacy of pragmatic coping in accounts of the mind and the organism-relativity of experience. In Section 5, I argue that this pragmatic, “narcissistic” (Akins 1996) character of prediction error minimization undermines its ability to explain the distinctive *normativity* of intentionality. Finally, in Section 6 I argue that predictive processing therefore positively *mandates* an extra-neural account of intentional content of exactly the sort that pragmatism’s communitarian vision of human thought—the third and most controversial pragmatist commitment I outline in Section 2—can provide. I then offer a preliminary sketch of what such an account might look like.

First, however, I address an obvious question: why bring these two frameworks into contact with one another? Why relate a contemporary research programme in cognitive neuroscience to a *philosophical* tradition that emerged in late nineteenth century North America? I think there are three important reasons.

First, pragmatism is and always has been a naturalistic philosophical movement, “stressing the continuity of human beings with the rest of nature” and emphasising “the need for philosophy to be informed by, and open to, the significance of novel scientific developments” (Bernstein 2010, 8; cf. also Bacon 2012). Of course, the naturalism in question is rarely *metaphysical naturalism*,¹ the thesis that reality is exclusively and exhaustively described in the vocabulary of natural science, but rather what Price (2013) helpfully terms “subject naturalism,” the view that philosophy should start from the assumption that we are natural creatures in a natural environment (Brandom 2010, 208–10). From the perspective of pragmatists, then, it is interesting to see how an extremely ambitious theory like predictive processing bears on its central commitments. After all, it aspires to tell us what kinds of natural creatures we are.

¹ A notable exception to this claim is Quine (e.g. 1960), although Quine’s relation to the pragmatist tradition is complicated (cf. Hylton 2007).

This leads to a second reason, however: as noted above, insofar as predictive processing *does* bear on the pragmatist tradition, it seems to be in explicit tension with it. Peirce first introduced pragmatism at the Metaphysical Club in late nineteenth century Massachusetts with the aim of rejecting “the spirit of Cartesianism” (Peirce 2017), and a unifying aspiration among pragmatist thinkers has been to disarm philosophy from a picture of minds as “mirrors of nature” (Rorty 1979), passively re-presenting the nature and structure of some pre-given reality (Bacon 2012; Price 2011). For this reason, the early American pragmatists are often seen as helping to give birth to the tradition of anti-representationalism in *psychology* whose later manifestations were mid-twentieth century scientific behaviourism and the tradition of ecological psychology (Chemero 2009, ch.2; also Heft 2001; Rockwell 2005). Predictive processing, by contrast, at least appears to be a robustly representational theory of the mind, advancing what Hohwy (2013) calls a neuroscientific vindication of a conception of minds as mirrors of nature and a neurobiological approach to intentionality (see Section 2.3 below). Hostility to such a Cartesian framework has thus motivated hostility to predictive processing itself (cf. Anderson and Chemero 2013).

Several authors (especially Clark 2016; but see also Allen and Friston 2016; Anderson 2017; Bruineberg et al. 2016; Gallagher and Allen 2016; Hutto 2017; Kirchhoff 2017; Seth 2015) have recently sought to undermine these putative implications of the theory. Their focus has consisted predominantly in connecting predictive processing to the explicit research agenda and implications of “4E cognition,” however.² Whilst I draw on some of this excellent work below, my aim here is to address broader themes in the pragmatist tradition that can be articulated without the idiosyncratic vocabulary of that research programme and that are not shackled to some of its more “radical” commitments (see Section 5). As I will present it, pragmatism provides an extremely general conception of the human animal within which more specific philosophical and scientific research can be conducted.

Finally, this paper is not just defensive but *constructive*: as I argue below, prominent and misguided Cartesian interpretations of predictive processing have obscured the way in which the two frameworks can be brought into mutually illuminating contact with one another. Substantiating this claim is the aim of Sections 4, 5, and 6.

First, however, I introduce what the defining commitments in pragmatism *are* (Section 2), before turning to outline those tenets of predictive processing under its maximally Cartesian guise that contradict each of these commitments (Section 3).

2 Pragmatism, mind, and experience

It would be misguided to try and specify a set of specific theoretical convictions shared by all pragmatists (Bacon 2012, i). My aim here is thus more modest. First, I will focus on commitments in the pragmatist tradition that are sufficiently *general* to encompass what Godfrey-Smith (2015, 807) calls an otherwise “diverse and mutable collection of ideas” within a single, overarching framework. The hope is that this generality will enable the specific views and theories of pragmatist thinkers to be viewed as *manifestations* of underlying, more general commitments, and thus allow for a conception of

² “4e” for embodied, embedded, enactive, and extended cognition (cf. Clark 2011).

pragmatism as an *evolving* tradition not too strongly shackled to the peculiarities of any *individual* pragmatist's ideas. Second, I will restrict my scope to those general commitments that explicitly concern the nature of thought, experience, and the mind.

Still, I stress that the following treatment is *not* intended to be either *exhaustive* or *definitive*, and it is *not* intended to be in competition with treatments that identify *different* core themes in the pragmatist tradition. To take one important example, Menary (2015) has recently conducted an excellent survey of the pragmatist foundations of the “pragmatic turn” in cognitive science, and explicitly related these foundations to predictive processing. The three core themes that he outlines, however, are distinct from the ones I have chosen to focus on. Nevertheless, my treatment is intended to complement Menary's, not oppose it. I hope that by focusing on alternative themes, the conversation he instigated can be extended and broadened.

Despite these qualifications, I do think that *in conjunction with one another* the three commitments I articulate constitute something like a centre of theoretical gravity in the pragmatist tradition, an attractor around which otherwise heterogeneous interests and substantive positions have arisen. This section articulates and elaborates those commitments.

2.1 Coping, not copying

The first commitment is to the *functional primacy of action and pragmatic success in accounts of the mind*. As Brandom (2002a, 40) puts it, pragmatism is “a movement *centred* on the primacy of the practical” (my emphasis). On this view, we should understand thought and language not as passive mediums for re-presenting the nature of nature but as practical tools for facilitating action and more general kinds of practical success. In the Deweyan slogan popularised by Rorty (1989, 1999), cognition is for “coping, not copying.”

This conviction constituted a unifying theme in the early American pragmatist tradition. Peirce, James and Dewey all sought to understand thought in terms of its role in guiding *action* (Godfrey-Smith 2015). Peirce et al. (1974) argued that philosophy should examine the nature of thought and ideas in terms of their difference to human behaviour, and his “pragmatic maxim” ties the meanings of concepts to their practical effects. James (1907/2000) famously extended this maxim beyond the realm of empirical enquiry to account for the significance of all areas of human thought and practice, not just science. And Dewey (1925, 1948) developed elaborate proposals of the relationship between theoretical and practical goals in a sustained effort to undermine what he called the “spectator theory of knowledge,” a theory in which cognitive processes are decoupled from practical ends (cf. Godfrey-Smith 2015, 804). “Concepts, theories and systems of thought... are tools,” he argued. “As in the case of all tools their value resides not in themselves but in their capacity to work shown in the consequences of their use” (Dewey 1920/1948, 145).

The emphasis on the relationship between belief and *action* is less pronounced in the work of neo-pragmatist authors like Quine, Rorty, Brandom, and Price (Godfrey-Smith 2015). Nevertheless, the primacy of *pragmatic coping* more generally in understanding thought and language has remained central. Rorty (1989, 1999), for example, (in)famously argues that numerous philosophical problems arise from a misguided “representationalism” that fails to take account of our status as the biological upshots of

fundamentally *Darwinian* processes, biological upshots whose concern is with coping with the world, not copying it. Further, this commitment to the “primacy of the practical” is continued in the work of prominent neo-pragmatists like Brandom (1994, 2010) and Price (2011, 2013), both of whom seek to understand our concepts and linguistic frameworks not first and foremost as *representations* of reality but as tools that serve various (often orthogonal) *practical functions* in our lives.

What unifies these diverse philosophers is a shared conviction that the mind should be understood not as an inner arena designed to recapitulate reality, but as a functionally crucial *node* in a complex web of forces that enable an organism to cope with its variegated environments given its many practical ends. Viewed from this general perspective, we can see the more specific views of individual pragmatist authors—Peirce’s pragmatic maxim, for example, or James’s theory of truth—as *manifestations* of this general stance, not essential components of it.

A common objection to an emphasis on *coping* over *copying* is that the two are not mutually exclusive—indeed, that copying seems to be effectively mandated by the kinds of practical coping exhibited by humans and other animals (Blackburn 2006; see Section 5.2 below). To see why pragmatists generally reject this line of reasoning, we must turn to a second core theme in the tradition.³

2.2 Contingency

If action and pragmatic coping are the most fundamental concepts in the pragmatist tradition, *contingency* is not far behind (Price 2011; Rorty 1989). In its hostility to the metaphors of “mirroring” and “copying” resides a profound attachment to the importance of the *subject* in the construction of its experienced reality. On this view, contingent properties of the organism are functionally crucial to the contents of its experience, thereby undermining a common dualism in which the contents of mind reflect independently identifiable contents in the world. As Dewey (LW, 14, 17) put it, “the organism – the self, the “subject” of action – is a *factor within experience* and not something outside of it to which experiences are attached as the self’s private property” (my emphasis).⁴

Whilst less central to Peirce’s work, this commitment to the constructive nature of experience lay at the core of James’s pragmatist vision, underlying his famous remark that “the trail of the human serpent is... over everything” (James 1907/2000, 68). For James, our status as a certain kind of *creature* inextricably colours our commerce with the world. Likewise, Dewey’s famous interactive conception of knowledge holds that knowledge of the world is formed as an adaptive *response* to environmental circumstances given the agent’s needs and purposes, an ongoing process in which the subject moulds and constructs the very environments it inhabits (Dewey 1925; Godfrey-Smith 2013). For this reason, Dewey was a central influence on Gibson and the tradition of

³ As I argue below, the existence of internal representations is consistent with pragmatism. Peirce, for example, outlined a sophisticated analysis of representation and was himself committed to mental representations. Pragmatists are rather against a particular *understanding* of what representation amounts to (Price 2013). I argue below that what they reject is twofold: first, an understanding of internal representations as “mirrors of nature” (see S3); and second, the idea that *intentionality* (the paradigm of which is propositional thought) can exist independently of sociolinguistic practices (see S5).

⁴ Quoted in Bacon (2012, 52).

ecological psychology, where the idea that an organism's perceived environment is fundamentally a world of "affordances"—roughly, opportunities for environmental intervention (Chemero 2009)—highlights the functional importance of its practical interests, abilities and morphology in bringing forth its experienced world (Gibson 1979).⁵

A similar conviction is central to the work of various neo-pragmatists. For Rorty, for example, the contents of human thought and language can only be understood in the context of practices with specific *ends* that implicate such contingent features as our evolutionary ancestry, biological endowment, cognitive structure, spatiotemporal location and sociocultural context—and, above all, our culturally mediated interests and purposes given this variegated background. Likewise, Price's (2013) "global expressivism" seeks to understand our various vocabularies in terms of the "contingent, shared dispositions" and "practical stances" they exhibit, rather than in terms of representational relations like *truth* or *reference* (Price 2013, 62). Indeed, as with Quine (1960) and Rorty (1989), Price (2013, 63) argues that there are "no cases in which the contingencies on the speakers' side go to zero," in line with James's suggestion about the trail of the human serpent.

The functional primacy of pragmatic success is therefore *relativized* to the contingent properties and interests of the subject pursuing that success. In this way, the organism—in the case of human beings, a product of a complex web of biological and cultural circumstance—makes an active contribution to the construction of its experienced world. For this reason, Price (2011, 30) recommends that we abandon the metaphor of a *mirror* in our conception of mind and language in favour of a *holographic data projector*, a metaphor intended to capture the extent to which an organism's experienced world is as much a matter of dynamic *creation* as *reflection*. The "projection" here constitutes the organism's "manifest image" (Sellars 1963) or "narcissistic ontology" (Dennett 2013, 71), as much conditioned by the properties and interests it finds itself *with* as by the environment it finds itself *in*.

2.3 The social mind

The foregoing commitments might leave the impression that pragmatism is an *individualist* creed, a movement committed to a fundamentally asocial mind. Of course, nothing could be further from the truth. Indeed, some have argued that an emphasis on the philosophical significance of *social practices* is the core theme in the pragmatist tradition (Bacon 2012). The third commitment I will focus on, then, is to the idea that distinctively human forms of thought are made possible only within the context of normatively structured social practices among language-users like ourselves. In a slogan: intentionality—the very contents of our thoughts and linguistic expressions—is an inextricably *social* phenomenon.

This commitment to a "communitarian" or "social approach to intentionality" (Satne 2017, 528) became central to the pragmatist tradition with the work of Dewey, who advanced a "social theory of mind" in which "genuine thought is made possible only within the context of a language-using community" (Godfrey-Smith 2015, 805). For Dewey, *thought* arises within the context of symbol-using behaviour, and symbol-using

⁵ This commitment has strongly influenced the enactivist tradition (Varela et al. 1991; Thompson 2007).

behaviour is a distinctively *social* activity, even if one can subsequently exploit this socially endowed ability internally in the form of inner speech (Dewey 1925; cf. also Godfrey-Smith 2013, 287). “Language,” he wrote, “is the cherishing mother of all significance” (Dewey 1925, 154). A similar conviction surfaced in Mead’s work, another early pragmatist thinker who “developed a comprehensive social theory of action and language” (Bernstein 2010, 8). Indeed, these authors’ emphasis on the sociolinguistic nature of human thought prefigured its emergence in the work of later philosophers such as Wittgenstein, Quine, Rorty and Brandom by some decades.

Among contemporary pragmatists, Brandom (1994, 2010) is the most prominent advocate of this view, although it also features prominently in the work of Haugeland (1990) (who self-identifies as a neo-pragmatist (Haugeland 1990, 422)) and in Price’s (2011, ch.8) hypothesis that the *truth* norm emerges only in the context of social disagreement—an interesting suggestion I return to in Section 6. In Brandom’s (1994) extended treatment of the idea, the semantic norms that underlie intentional content arise within the social practice of “giving and asking for reasons” (Sellars 1963)—that is, within structured social interactions in which assertions are held answerable to socially instituted inferential norms. The distinctively *human* mind and the unique kinds of normative statuses it can exhibit are thus construction of *culture* (Brandom 2002a, 2002b; cf. also Haugeland 1990). The upshot is that “we can only make sense of contentful thinking in the context of shared ways of life in which social norm compliance is developed, maintained and stabilized through social practices” (Hutto and Satne 2015, 527).

For this reason, this third pragmatist commitment is diametrically opposed to a more traditional and orthodox conception of intentionality, in which the contents of our thoughts exist *prior to* and are simply *communicated with* language, understood as “a set of publicly accessible signs that are combined according to certain rules to form meaningful sentences” (Satne 2017, 529; cf. Grice 1957; Lewis 1975). On this view, “original (i.e. *underived*) intentionality” (Searle 1980) resides first in the head, and the cognitive significance of cultural artifacts like natural languages is *derived*. For pragmatists, by contrast:

“The idea is that contentful tokens, like ritual objects, customary performances, and tools, occupy determinate niches within the social fabric – and these niches “define” them as what they are. Only in virtue of such culturally instituted roles can tokens have contents at all” (Haugeland 1990, 404).

On this view, then, the *human* mind is a constitutively social mind, not formed *prior to* the human animal’s immersion in sociolinguistic practices but rather emergent *from* such practices. Our third pragmatist commitment is thus maximally un-Cartesian: the very contents of our thoughts are a function of our immersion in social practices of evaluation within which assertions and judgments are held answerable to culturally instituted norms.

2.4 Summary

This whistle-stop tour of the pragmatist tradition has been skeletal. Nevertheless, it presents a very *general* picture that I think has been largely common to the tradition, a

picture that has grounded the many more concrete commitments and theses of specific pragmatist authors. This is a picture in which minds function in the service of pragmatic coping, in which contingent and idiosyncratic contributions of the organism in its biological and social milieu are functionally crucial to the construction of its experienced world, and in which distinctively human forms of thought are made possible only within the nexus of normatively structured sociolinguistic interactions characteristic of human life.

My thesis is that this extremely broad vision of the human animal meshes beautifully with recent work from cognitive neuroscience, where a “new theory is taking hold” (Hohwy 2013, 1), a “paradigm shift” (Friston et al. 2017, 1) that some believe is ushering us into a new, “predictive era” in the history of psychology (Gładziejewski 2015). Specifically, I contend that this emerging theory powerfully *vindicates* and *illuminates* the first two commitments enumerated above, and positively *mandates* the third.

First, however, I turn to an elegant presentation and interpretation of this emerging framework in cognitive neuroscience that draws exactly the opposite conclusion.

3 The Cartesian predictive mind

The term “predictive processing” is used in numerous ways in both the scientific and philosophical literature. As I use the term (following Clark 2016), it refers to a maximally ambitious theoretical framework for understanding the brain as an “organ for prediction error minimization” (Hohwy 2014, 1), a process that implicates *hierarchical predictive coding* and *probabilistic generative models* of the body and environment’s causal structure (cf. Clark 2013, 2016; Friston 2010; Hohwy 2013; Kiefer and Hohwy 2017). This theoretical framework has generated an enormous amount of research, excitement, and hostility in recent years. In this section, I first outline those core tenets of the theory in its maximally “Cartesian” guise that can plausibly be extracted from the work of Hohwy (2013, 2014) and others (Kiefer and Hohwy 2017), and then explain why the resultant conception of cognition is so radically at odds with the pragmatist vision of mind outlined in Section 2.⁶ In later sections, I identify subtler aspects of the framework, along with its connection to the “free-energy principle” (Friston 2009, 2010).

First, then, predictive processing is often presented as offering a solution to the “problem of perception” (Hohwy 2013, ch.1). Following Helmholtz, the problem is how to veridically *infer* the distal causes of the brain’s sensory inputs given access only to the sensory inputs themselves. The *problem* is that such sensory evidence is inherently *noisy* and *ambiguous*, such that its environmental causes are radically *underdetermined*. *Bayesian perceptual psychology* models the solution to this problem in terms of Bayesian inference, where *Bayes’s theorem* identifies an optimal way of updating one’s beliefs given new evidence under conditions of uncertainty (Rescorla 2013). Specifically, it states that the posterior probability of a hypothesis given new

⁶ I don’t go so far as to positively ascribe the interpretation of predictive processing in this Section to Hohwy, if only because he is alert to many of the points raised in Section 4. Nevertheless, it *can* plausibly be reconstructed from his work (e.g. Hohwy 2013, 2014; Kiefer and Hohwy 2017).

evidence $P(H/E)$ is proportional to that hypothesis' *likelihood* $P(E/H)$ —how well the hypothesis *predicts* the evidence—weighted by its *prior probability* $P(H)$:

$$\left(\textit{Baye's Theorem}\right) P(H|E) = P(E|H) * P(H) / P(E).$$

Despite the widespread influence of Bayesian modelling in perceptual psychology and cognitive science more generally, it confronts two big challenges (Rescorla 2013; Tenenbaum et al. 2011): first, how is Bayesian inference approximated *algorithmically* in the brain? When dealing with sufficiently large hypothesis spaces, Bayesian inference can be extremely slow and often computationally intractable, which has led researchers in statistics and machine learning to develop various procedures that *approximate* the results of exact Bayesian inference (Penny 2012). Second, how are such algorithms implemented in the brain's neural networks?

Predictive processing can then be seen—and is often presented—as an answer to these questions, explaining how brains come to approximate Bayesian inference in cortical information-processing through *hierarchical predictive coding* and *precision-weighted prediction error minimization* (Clark 2016; Hohwy 2013; Penny 2012).

To understand how this works, consider first how Bayesian inference can be operationalised in terms of *predictions* and *precision-weighted prediction errors*. The idea is relatively straightforward: if one assumes Gaussian probability distributions, one can calculate one's *posterior* distribution by comparing the mean value p of the prior distribution with the mean value e of the evidence to compute a mismatch signal or *prediction error* (Denève and Jardri 2016).⁷ Bayes' theorem then dictates how to *weight* the prediction error in updating the prior, which in turn determines the “learning rate”: the higher the weight assigned to the prediction error, the more the agent learns from the evidence, and thus the more it should update its priors (Hohwy 2017). This weight should thus be evaluated by comparing the uncertainty of each source of information, which can be calculated by the relative *precisions* (the inverse of the *variance*) of the two distributions. Intuitively, if one's expectations or evidence are extremely *noisy* (i.e. highly variable) they should not influence the posterior estimate as much. Likewise, as an agent *learns more*, its priors should become increasingly *precise*, thus ensuring that its prior knowledge plays a greater role in guiding inference.

For this to work in the case of real-world perception, however, simple Bayesian inference is not enough. The easiest way to see this is as follows: if a system performs Bayesian inference, it will keep prediction error at an optimal minimum, limited by ineliminable noise (Hohwy 2017). (This simply reflects the optimal character of Bayesian inference). The *sensory evidence*, however, is a function of complex, interacting hidden causes in a dynamic environment that contains regularities nested within regularities. For a system to optimally minimize prediction error under such conditions, then, it requires some means of *deconvolving* the structural elements of such environments to separate out regularities at different scales. This requires *hierarchical Bayesian inference*: if sensory input is a function of a complex, hierarchically nested causal structure, the perceptual system must effectively *invert* this hierarchical

⁷ For convenience, I run together the distinction between distributions and density functions throughout, as it is irrelevant to the paper.

structure, forming expectations over different regularities at different scales and exploiting such longer-term expectations to modulate its *learning rate* in context-sensitive ways (cf. Clark 2016; Hohwy 2017). The resultant inferential architecture functions as what's known as a “hierarchical Gaussian filter,” filtering regularities in the evidence at multiple scales and exploiting such hierarchically distributed expectations to predict the incoming evidence in context-sensitive ways (Mathys et al. 2014). In such an architecture, the hypotheses at each level then function as the sensory evidence for the level above (Denève and Jardri 2016).

Predictive processing then advances a “process theory” for *how* this process of hierarchical Bayesian inference is approximated and implemented in the brain (Friston et al. 2017). Recall that a system that follows *exact* Bayesian inference in the manner just described will optimally minimize long-term prediction error. If this is right, however, then—subject to certain technical assumptions that fall beyond the scope of the current paper—a system can be expected to *approximate* the results of exact Bayesian inference in proportion to its ability to minimize long-term prediction error (Kiefer and Hohwy 2017). The importance of this observation is this: whilst it is implausible to think that the brain explicitly follows exact Bayesian inference in the manner described above, it is not implausible to suppose that the brain can construct hierarchically structured models of the world with which to anticipate the sensory signal via recurrent feedback connections (Hohwy 2017; Kiefer and Hohwy 2017). Indeed, there is extensive evidence of both this hierarchical structure in the neocortex and the prolific role of “backwards” or “top-down” connections carrying signals back towards primary sensory areas (Pendl et al. 2017).

Predictive processing capitalizes on such evidence. At its core is the idea of *predictive coding*, a data compression strategy whereby only the unpredicted elements of a signal—the *prediction errors*—are fed forward for further stages of information-processing. In the case of the brain, the idea is that “top-down” synaptic connections in the brain from higher (e.g. frontal or temporal) cortical areas carry predictions of activity to lower levels in the cortical hierarchy until they reach proximal sensory input, and “bottom-up” connections carry the residual *prediction errors* (Clark 2013).⁸ When this simple information-processing strategy is combined with a vision of the brain as an organ whose function is to minimize prediction errors, the upshot is a process that—as Hohwy (2017, 10) puts it—is “more mechanistic and less literally inferential.” Roughly, the proposal is that as the brain becomes increasingly successful at minimizing prediction error, it comes to both *install* and then *continually update* a veridical generative model of environmental causes—a “mirror of nature” (Hohwy 2013)—in a way that *approximates* the results of exact Bayesian inference in the long run. As Hohwy (2013, 55) puts it, “neuronal populations are just trying to generate activity that anticipates their input. In the process of doing this they realize Bayesian inference.”

That—in an extremely skeletal nutshell—is predictive processing’s account of learning and “perceptual inference.” Of course, the foregoing overview only scratches the surface of the relevant technical literature, and says nothing about how other cognitive phenomena are supposed to fit in within this framework (see Section 4 below). Nevertheless, this summary should give some indication of the extreme *elegance* of

⁸ Given space constraints, I ignore the role of lateral connections in this brief overview (cf. Kiefer and Hohwy 2017).

this account—specifically, its beautiful marriage of tradition Cartesian internalism with the technical resources of contemporary machine learning, statistics, and cognitive neuroscience. Encased within a dark skull and without supervision from an external teacher, predictive brains exploit the statistical patterns in their sensory input to reconstruct the causal-probabilistic structure of the objective reality beyond it—a “virtual reality” (Hobson et al. 2014) or “controlled hallucination” (Grush 2004) with which they can *form judgements, tests hypotheses, and revise beliefs*. From an initial Helmholtzian vision of perception as abductive inference, we arrive at an explanation of how the optimality of Bayesian inference might genuinely be *approximated* and *implemented* in cortical circuitry via this recursive process of prediction revision and model-updating.

More importantly, however, this brief overview should exemplify how profoundly this vision of the mind contradicts each of the defining pragmatist commitments outlined in Section 2. In place of the primacy of the practical (Section 2.1) and the profound ways in which the organism structures its experienced world (Section 2.2), the brain is viewed as an organ for veridically reconstructing the objective structure of the signal source via an approximation to optimal Bayesian inference. In addition, this process confers on the brain *intentional states* (Section 2.3), enabling it to form truth-conditional hypotheses about the causes of its sensory input and continually test these hypotheses against sensory evidence—a brain-bound conception of original intentionality radically at odds with the *social* theory of mind advanced in the pragmatist tradition (Hohwy 2013; Kiefer and Hohwy 2017).

It is difficult to imagine a conception of mentality more in tension with the pragmatist themes enumerated in Section 2. It is also—or so I argue next—mistaken as an interpretation of the science. Specifically, it rests on a dubious philosophical overlay not motivated by the content of predictive processing itself.

4 The pragmatic predictive brain

I began above with the “problem of perception”: given access only to the evolving activity at the organism’s sensory transducers, how can the brain infer the objective structure of the ambient environment? This “problem,” of course, perfectly exemplifies what Dewey (1925, 1948) referred to as the “spectator theory of knowledge”: encased within a lonely skull, the brain is viewed as a statistical inference engine analogous to the *artificial* neural networks familiar from machine learning—that is, systems explicitly and intentionally designed by human beings to exploit the statistical patterns in their input data to recover the objective structure responsible for that data (Kiefer and Hohwy 2017). On this view, the function of prediction error minimization is to enable brains to produce a copy—an internal mirror—of the objective structure of the ambient environment.

I argued in Section 2 that pragmatism is fundamentally opposed to any such vision of cognitive activity. In this section, I argue that predictive processing is as well (cf. Anderson 2017; Barrett 2016; Bruineberg et al. 2016; Clark 2016; Seth 2015; Williams 2017).⁹ To see this, one must first relate predictive processing to the “free-

⁹ Anderson (2017) chastises the “problem of perception” for assuming that sensory input is *ambiguous*, an issue I do not address in this paper. In addition, Anderson advocates a regulative perspective on brain function as an alternative to the problem of perception, but fails to note that this perspective is *endorsed* by predictive processing itself (cf. Williams 2017).

energy principle” as advanced by Friston (2009, 2010). Any thorough overview of the free-energy principle falls beyond the scope of the current paper. Nevertheless, its central ideas can be gleaned relatively straightforwardly, and it is crucial in adjudicating the debate I address here. Specifically, it reveals the underlying *rationale* for prediction error minimization as a special case of a deeper imperative in biological agents to self-organize around their homeostatic set-points under conditions tending towards increasing disorder (Friston 2010). That is, it reveals prediction error minimization as the solution to a fundamentally *pragmatic*—not *representational*—problem (Bruineberg et al. 2016; Clark 2016; Seth 2015). As I argue in later sections, this has enormous significance for discussions about intentionality and the predictive brain.

The free-energy principle begins from the observation that “the defining characteristic of biological systems is that they maintain their states and form in the face of a constantly changing environment,” thereby somehow violating “the fluctuation theorem, which generalizes the second law of thermodynamics” (Friston 2010, 1). Given this surprising characteristic, one can derive the following fundamental job description for brains: “to regulate the organism’s internal milieu” (Sterling and Laughlin 2015, xvi.). This perspective on brain function effectively follows insights from mid-twentieth-century cybernetics in modelling brains as *regulators* responsible for maintaining essential homeostatic variables within viable bounds (Conant and Ashby 1970; cf. Seth 2015; Williams and Colling 2017). The novel contribution of the free-energy principle is to provide an information-theoretic interpretation both of what this homeostatic process amounts to and how it can be achieved in biological agents via the minimization of a quantity to which they have internal access. Specifically, it states “that any self-organizing system that is at equilibrium with its environment must minimize its free energy” (Friston 2010, 1).

For our purposes, there are two important components to this idea. The first is the idea that homeostasis can be described as the minimization of long-term, average *surprisal*, where “surprisal” is an information-theoretic quantity that names the improbability of an outcome relative to a probability distribution (Friston 2010; cf. also Hohwy 2015). The intuitive idea here is straightforward. If a probability distribution is defined over possible states of an organism, *homeostasis* (i.e. *being alive*) requires it to occupy and revisit an extremely narrow subset of such possible states. Relative to optimal homeostatic regulation, then, deviations from homeostasis effectively result in increasingly improbable (highly “surprising”) states. As such, homeostasis can be *described* as the minimization of long-term surprisal. As many have noted, this makes surprisal profoundly *organism-relative*: it names states that are non-optimal *relative to an organism’s contingent phenotype* (Friston 2010, 2; cf. also Bruineberg et al. 2016; Seth 2015; Williams 2017).

The second important idea is a proposal concerning *how* biological systems can avoid surprising states. Although the technical details are formidable here, the central point for our concerns is Friston’s proposal that a tractable optimization task for brains to perform that *approximates* the minimization of long-term surprisal is the minimization of *variational free energy*, another information-theoretic quantity that provides an upper bound on surprisal and—under a set of technical assumptions (cf. Friston 2009, 2010; Hohwy 2015)—*translates to long-term, average prediction error* (Friston 2010). In other words, the *problem* that prediction error minimization solves is homeostatic

regulation, such that the former “is, essentially, a tool for self-organisation” (Gładziejewski 2015, 563).

If this is right, the upshot is simple: the *function* of prediction error minimization is *not* the veridical reconstruction of the organism’s ambient environment, but homeostatic regulation—a problem that is positively *defined* in terms of the organism’s idiosyncratic practical interests. As Barrett (2016, 3) puts it:

“A brain did not evolve for rationality... or accurate perception. All brains accomplish the same core task (Sterling and Laughlin 2015): to efficiently ensure resources for physiological systems within an animal’s body... so that an animal can grow, survive and reproduce.”

Further, note that the process described in Section 3 of updating top-down predictions to effectively anticipate the incoming signal is in itself *impotent* with respect to this goal. That is, it is no good predicting sensory inputs if those inputs signify the organism’s death (Hohwy 2013, 85). What the brain needs, then, is some means of actively *changing* its sensory inputs. Within the context of predictive processing, this corresponds to *active inference*, the opposite of “perceptual inference” as outlined in Section 3: rather than updating top-down predictions to bring them into alignment with sensory inputs, the brain actively moves the body to generate the sensory inputs it has been designed to expect (Friston 2010; Hohwy 2013, 84–9). Ultimately, the fundamental predictions here concern the defining homeostatic variables of the organism (Friston 2010). As Bruineberg et al. (2016) stress, this makes *active inference* functionally primary in predictive processing: it is only by acting on the environment to maintain its optimal states that an organism can *change* its sensory input and thus avoid surprising experiences. Any “perceptual inference” that occurs is thus answerable to this pragmatic end (see Section 5 below).

These observations have two important implications. First, the superficial appearance of a dramatic tension between predictive processing and the first pragmatist commitment outlined in Section 2 dissolves to reveal something much more interesting. When situated in this broader theoretical context, predictive processing does not just *vindicate* but powerfully *illuminates* that commitment: the schematic view that pragmatic coping should be functionally primary in accounts of the mind is given concrete expression in a scientific vision of how cognitive processes emerge in the service of the ultimate kind of pragmatic success—namely, maintaining the organism’s viability under conditions described by the second law of thermodynamics.

Second, this framework reveals how predictive processing might further illuminate the *second* commitment outlined in Section 2: once we situate prediction error minimization in this pragmatic context of homeostatic regulation, it becomes clear that any vision of the brain as simply recovering the objective structure of the distal environment becomes deeply problematic (Anderson 2017; Bruineberg et al. 2016; Clark 2016; Williams 2017). Instead, we must confront the profound *organism-relativity of experience*. In other words, we should expect profoundly *narcissistic nervous systems* (cf. Akins 196; Williams 2017): that is, nervous systems whose responses to the environment are mediated entirely by the pragmatic, homeostatic concerns of the organism.¹⁰

¹⁰ This generalises Akins’ (1996) famous idea of “narcissistic sensory systems.”

Barrett (2016, 2017) introduces the helpful concept of the “affective niche” to capture this organism-relative character of the brains’ generative model within the predictive processing framework. The affective niche is “species-specific” and includes “only the parts of the animal’s physical surroundings that the brain has judged relevant for growth, survival and reproduction” (Barrett 2016, 6).

These considerations, then, suggest that the initial appearance of a deep conflict between pragmatism and predictive processing is illusory. Far from an image of minds as passive spectators on the world, recovering the objective structure of the environment like an idealised scientist, predictive processing advances a fundamentally pragmatic brain, striving to maintain the viability of the organism under hostile conditions and in so doing actively *generating* an affective niche—an experienced world structured by the idiosyncratic practical interests of the organism. What emerges is something much closer to Price’s (2011) metaphor of a “holographic data projector” (see Section 2.2) than a passive reflection of an independently identifiable world. As Clark (2015, 4) puts it, it is a vision of experience that is “maximally distant from a passive (“mirrors of nature”) story.”

Nevertheless, one might still object that predictive processing’s apparent consilience with the first two pragmatist themes outlined in Section 2 is irrelevant given its stark opposition to the third. After all, even if predictive brains *are* pragmatic and narcissistic, don’t they still generate in-the-head intentional states in a manner flatly inconsistent with pragmatism’s social vision of mind? Indeed, even many authors who *agree* with the points I advance in this section nevertheless accept that predictive processing *does* explain the emergence of *original intentionality* in brain states (e.g. Barrett 2016; Clark 2016; Seth 2015).

I next argue that this is mistaken. In fact, once one situates prediction error minimization in the broader theoretical context revealed in this section, it becomes clear that it lacks the requisite *normativity* to explain intentionality. The upshot of this—or so I argue in Section 6—is that the direction of explanatory illumination runs *in both directions*: predictive processing positively *mandates* an extra-neural account of intentional states of just the sort that pragmatism’s *communitarian* vision of intentional content—the *third* pragmatist theme outlined in Section 2—can provide.

5 Intentionality and the predictive mind

In this section, I argue that prediction error minimization is inadequate on its own to explain the emergence of original intentionality. I then consider and reject two objections to this argument. First, however, I raise an issue that must be set aside.

Consonant with a broader debate in recent cognitive science, there is a growing literature on to what extent predictive processing advances a *representational* theory of cognition, with some arguing that it does (Clark 2016; Gładziejewski 2015; Hohwy 2013; Williams 2017; Williams and Colling 2017) and others that it doesn’t (Gallagher and Allen 2016; Bruineberg et al. 2016; Hutto 2017). In line with common assumptions in the philosophical literature, this debate is often *identified* with the debate concerning to what extent it explains *intentionality*, typically understood in terms of *content* and thus *veridicality conditions*.

I regard this as a mistake (see Williams 2017). Specifically, it runs together two distinct *species* of representation: first, the concept of representation as *proxy* or *stand-*

in; and second, the concept of representation as *judgement*, where “judgement” subsumes any representational state identified with *veridicality conditions* and thus includes the essential *normativity* of intentionality (see below). Of course, internal proxies for environmental states might play a central role in the *explanation* of intentionality (see Section 6). The point is just that they are not identical with intentional states.

There are three reasons why. First, the exploitation of structures that function as proxies for other states or conditions is plausibly *ubiquitous* in biological systems (Bechtel 2008, ch.5; Williams and Colling 2017). Indeed, it is a central tenet of the “good regulator theorem” (Conant and Ashby 1970) developed by the free-energy principle that *all* optimal regulative systems exploit structures that are isomorphic to the systems they regulate (Friston 2010).¹¹ By contrast, it is not plausible that truth-evaluable intentional states are similarly ubiquitous.¹² Second, and relatedly, one can explain how a structure performs the role of a proxy without talking about content, veridicality conditions, or even the possibility of misrepresentation (cf. Bechtel 2008, ch.5). Intentional content is thus not *necessary* for something to function as a proxy. Finally, functioning as a proxy is not *sufficient* to possess intentional content: one cannot answer one of the hardest problems in the history of philosophy—namely, the emergence of intentionality—by pointing to the regulative behaviour of plants (Calvo and Friston 2017) (see Section 5.2 below).

To see the importance of this distinction, consider Anderson and Chemero’s (2013) claim that predictive processing conflates two distinct senses of the term “prediction.” On one sense of that word, “prediction1,” it identifies the local anticipatory role of signals within the cortical hierarchy as they are matched against activity at lower levels. On another sense, “prediction2,” it describes a *belief about how the world is*. Their contention is that all the relevant neuroscience gives you is prediction1, not prediction2, and thus that predictive processing is *non-representational*.

That is too quick, however. I agree with Anderson and Chemero (2013) that predictive processing is not entitled to prediction2 (see below). In other words, it does not explain original intentionality. Nevertheless, there is an important third option that they neglect (cf. Williams 2017): namely, that what explains the brain’s ability to successfully predict1 its sensory inputs and guide behaviour is that it exploits a proxy—the brain’s generative model—for interest-relative aspects of the environment’s causal structure. In *that* sense, then, predictive processing plausibly is representational. Indeed, plausibly *all* biological systems are. But it doesn’t follow that it explains *intentionality*.¹³ As such, it is not the relevant kind of “representation” that has concerned pragmatists (Rorty 1979 makes a similar point). To put the point bluntly: pragmatism is not undermined by thermostats—a paradigmatic regulative system that

¹¹ Of course, *which* systems they regulate is specified relative to their homeostatic function, ensuring consistency with Section 4.

¹² Sims (2016) argues otherwise, but this is because he simply takes for granted that predictive processing explains intentionality.

¹³ Of course, one can simply *stipulate* that veridicality conditions are the essential property of representations and thus seek to replace talk of the generative *model* with some other set of expressions (e.g. Bruineberg et al. 2016; Hutto 2017). This understanding of terms like “model” is at odds with how the term is understood in traditions such as cybernetics and control theory, however. Further, exploitable structural similarity is different from mere detection or causal covariation (see Williams and Colling 2017), so the trivialising arguments against such latter relations—advanced, for example, by Ramsey (2007)—do not work.

exploits a stand-in (the level of mercury) for an interest-relative (in this case, *our* interests) environmental variable (ambient temperature). It *would* be undermined if thermostats acquired truth-conditional beliefs via this regulative function (that is, independent of human interpretation).

With this cleared up, I turn now to explain why predictive processing cannot explain original intentionality.

5.1 Prediction error minimization and intentionality

For convenience, I will understand intentionality minimally as *truth conditions*, both because this is standard in the literature—for example, Kiefer and Hohwy (2017, 17) explicitly contend that predictive signals possess “full-fledged truth-evaluable content”—and because it characterises the paradigmatic intentional state of belief. My argument is straightforward: as noted in Section 4, the function of prediction error minimization is *not* representational success. Therefore there is a root mismatch between prediction error minimization and truth. Therefore it can’t explain intentionality.

To take this more slowly, intentionality is an inextricably *normative* phenomenon (Brandom 1994; Kripke 1982; McDowell 1994). That is, intentional states are essentially “liable to assessment of *correctness* of representation, which is a special way of being *answerable* or *responsible* to what is represented” (Brandom 1994, 6). Given this, an enormous amount of philosophical work has been devoted to explaining *how* intentional states acquire this normative status: what makes it the case that some alleged representational vehicle—an utterance, brain state, sentence, etc.—can be in *error* relative to how things are? In the current context, what makes it the case that some prediction1 is true if and only if some environmental condition obtains and false if it doesn’t?

An obvious suggestion—developed by Hohwy (2013, ch.8) at some length in an explicit attempt to address this issue of normativity—is that the relevant norm against which representational success is evaluated is *prediction error minimization*. That is, if brains are organs *for* prediction error minimization, can’t prediction1s be evaluated relative to their contribution to this overarching goal? On this view, “misperception... [is] perceptual inferences that move the creature away from this goal: misperception is when prediction error increases” (Hohwy 2013, 174). Hohwy (2013, 175-9) goes on to qualify this statement by replacing prediction error with *long-term, average prediction error* to guard against the possibility that prediction1s might on any given occasion minimize prediction error despite being false. Nevertheless, the basic idea remains the same: successful representation is identified with successful (long-term) prediction error minimization, such that we can identify the truth conditions of prediction1s by reference to those conditions in which they *function* properly. As such, the proposal bears a striking resemblance to teleosemantic approaches to intentional content in which representational error is explained by appeal to malfunctions in evolutionarily selected information-carrying structures (cf. Millikan 1984).¹⁴

¹⁴ Given space constraints, I ignore Kiefer and Hohwy’s (2017) recent proposal that one can explain misrepresentation in terms of the “Kullback-Leibler divergence.” The reason for this is twofold: first, this quantity measures the mismatch between two internal models (the generative and recognition model), and thus requires a prior story of how such model states acquire their contents, which I address in the text; and second, this mismatch effectively formalises the quantity of prediction error, so the same exact problems arise.

Given the argument of Section 4, the problem with this strategy should be evident. The *function* of prediction error minimization is *not* veridical representation of the world but homeostasis (Seth 2015). This function, we saw, is profoundly organism-relative: organisms find different states surprising, and so must minimize different quantities (cf. Friston 2010). Truth, however, is not organism-relative in this way. Specifically, this would flatly contradict the normativity of intentionality—namely, the fact that intentional states must be answerable to a genuinely *independent* standard against which they can be evaluated. Further, it would flatly contradict the inferential norms that underlie the application of the truth predicate: one cannot infer from the fact that something is useful for an organism—for example, from the perspective of biological homeostasis—that it is true (cf. Price 2011). As such, there is a root mismatch between prediction error minimization and representational success. Therefore the former cannot be used to explain the latter. Recall Barrett’s (2016, 3) point that predictive brains “did not evolve for rationality... or accurate perception.” As Hohwy (2013, 180) himself puts it, for an organism to minimize prediction error “is for it to organize itself in a far-from-equilibrium state, thereby insulating itself temporarily... from entropic disorder described by the second law of thermodynamics.” Optimal homeostatic regulation, however, is a different norm from *truth*. As such, it doesn’t provide the requisite normativity to ground intentional content. This argument effectively parallels what is probably the chief challenge to teleosemantics: namely, that there is similarly a root mismatch between biological function and representational success (Burge 2010).

There has recently been an emergence of “structural resemblance” theories of content within the predictive brain (Gładziejewski 2015; Williams 2017; Williams and Colling 2017), endorsed by Hohwy (Kiefer and Hohwy 2017) himself. The central idea here is that prediction error minimization induces a generative model that *mirrors* and so *resembles* the causal-probabilistic structure of the bodily and environmental causes of the brain’s evolving sensory inputs. Whilst Kiefer and Hohwy’s (2017) proposal is complex and nuanced, the central idea is relatively straightforward. They propose that the contents (i.e. truth-conditions) of predictive hypotheses are acquired through their place in the broader structure of the generative model from which they arise, which in turn recapitulates the causal structure of the environment in the manner just suggested. As such, the contents of such predictions can be *identified* with the possible worlds that correspond to (i.e. resemble) the state of the brain’s generative model.¹⁵ Such predictions are thus *true* just in case the *actual world* aligns with the possible world determined by the state of the model (Kiefer and Hohwy 2017, 24). In simplistic terms, one can think of this as comparing the state of a statistical model that veridically represents reality with the brain’s statistical model (Kiefer and Hohwy 2017, 24).¹⁶ In this way they

¹⁵ In more technical terms, the parameters of the generative model determine the *modal space* wherein possible worlds correspond to different (probability distributions over) the values of its random variables (cf. Kiefer and Hohwy 2017).

¹⁶ Kiefer and Hohwy (2017) follow Gładziejewski (2015) in distinguishing two distinct *kinds* of representation: roughly, the failure of the model to match the environment’s causal structure, and the falsity of judgements made within a model. My argument applies to each (because both require *failure*), but for convenience I focus on the latter case.

effectively advance an understanding of misrepresentation *not* tied to prediction error minimization: misrepresentation is simply a failure of correspondence.¹⁷

There are two problems with this proposal, however. The first concerns the similarity relationship. Given the *narcissistic* character of predictive brains, it is not obvious that the structure of the generative model *will* find corresponding structure in the world as described in some more objective (i.e. scientific) vocabulary (cf. Anderson 2017; Bruineberg et al. 2016). Just think of the many constituents of our “manifest image,” or what Dennett (2013) calls our “narcissistic ontology”: cuteness, sweetness, justice, beauty, and so on. Which parts of the *structure* of reality do they latch on to? Short of this mirroring relation, however, it is not obvious that the proposal makes sense.

This just reflects the second and deeper problem, however: appealing to structural resemblance does not address the relevant problem of normativity. Specifically, the question still arises how the relevant *norms* against which internal brain structures can be evaluated as true or false arise. The norm cannot be similarity itself (Cummins 1996): as Kiefer and Hohwy (2017, 22) themselves note, if the brain’s hierarchical Bayesian networks represent possible worlds *defined* by their respective structures, *misrepresentation* and so *intentionality* is impossible.¹⁸ One therefore needs “an *independent* standard of comparison to define misrepresentation” (Kiefer and Hohwy 2017, 22, fn.15). This, however, is the hard part: it is the challenge of explaining how the *judgements* arise that bring internal structures into normative comparison with the environment. And this returns us to the same problem: the relevant standard cannot be success at prediction error minimization because of the non-equivalence between this function and representational success.

I suspect that Kiefer and Hohwy (2017) are misled on this point by taking as their paradigm of a prediction error minimizing system an *artificial neural network*—that is, a thoroughly *passive* system explicitly *designed* by human agents to function as a representational device. Under such conditions, *we* can evaluate its internal states against its “target.” Biological agents, however, are nothing like this: as per the free-energy principle, their function is not to match predictions against sensory inputs passively received from the world, but to change the world to match their homeostatic expectations. As such, *their goal* is *not* truth. Therefore—contra Hohwy (2013)—one cannot appeal to this goal to explain intentionality.

5.2 Objections

Before I address the significance of this point for the current paper, I address two likely objections.

First, one might simply *deny* that there is the mismatch between prediction error minimization and representational success I have suggested. As far as I can see, there are two ways one might do this. First, one might concede that the *function* of prediction error minimization is not representational success, but argue that the latter is nevertheless necessary for the former. On this view, prediction error minimization *requires* the brain to objectively represent the world; as such, the former can stand in for

¹⁷ In fact, they argue that *both* are useful notions of misrepresentation.

¹⁸ Cummins (1996) addresses this problem by appeal to teleosemantics, but Kiefer and Hohwy (2017) explicitly reject this idea.

representational success. I think that *this* strategy is not attractive, largely for the reasons outlined in Section 4. Of course, brains must be sufficiently responsive to their environments to serve their pragmatic function of homeostatic regulation. But there is no reason to suppose that activities conducive to homeostatic regulation will produce *truth* if this responsiveness is entirely mediated through a pragmatic lens (c.f. Anderson 2017, 7–8).

A more interesting response contends that prediction error minimization *is* representational success—specifically, that it is success at representing *one's own existence*. As Friston (2010, 2) puts it, minimizing prediction error can be viewed as “maximizing the sensory evidence for the agent’s existence.” Whilst superficially attractive in the current context, the problem is that this is *the wrong kind of representational success*. If we identify minimizing prediction error with truth in this sense, we must acknowledge that truth is organism-relative, because surprisal is organism-relative, and *that* is the quantity that is being minimized in the predictive brain. But truth is not organism-relative in this way for the reasons enumerated above. Therefore this response cannot explain intentionality. Or—put another way—the challenge under discussion is to explain the distinctive *answerability* of intentional states. To minimize prediction error, however, is to effectively make the world answerable to *you*—to your expectations. It is thus the *wrong kind* of answerability.

Second, one might object that predictive processing and the free-energy principle are both inextricably *Bayesian* and pitched in terms of probability distributions.¹⁹ For example, we saw in Section 4 that the free-energy principle is cast in terms of minimizing surprisal, which can be cast in terms of Bayesian model optimization (Friston 2010). Relatedly, as noted in Section 3, minimizing prediction error is supposed to approximate the results of exact Bayesian inference over time (Hohwy 2017). As such, one might argue that such talk is inextricably *intentional*, and thus that one cannot accept these theoretical frameworks *without* intentionality. That is, *if* they are correct (the argument goes), they *must* explain intentionality.

I see two major problems with this objection.

First, insofar as these frameworks are Bayesian, it is Bayesian inference that plants and bacteria engage in (Calvo and Friston 2017). Of course, maybe plants and bacteria *do* have intentional states (cf. Sims 2016). The point is that this cannot be decided merely by pointing out that they satisfy a certain formal description. As is familiar from computer science, computational procedures and mathematical frameworks more generally can be *formally* described (i.e. in a way that abstracts away from *content*) (Egan 2013). Thus the mere fact that plants can be modelled with the formalism of probability theory does not settle whether they have intentional states. As Anderson (2017) points out, Bayes’s theorem *itself* reads the same whether it describes conditionalizing among scientists or parameter-setting in a control system.

Second, insofar as it is Bayesian inference, it is a very strange kind of Bayesian inference. Bayes’ theorem is typically thought of as a paradigm of rationality when one’s evidence is *given*. Under such conditions, conditionalizing in accordance with Bayes’s theorem will make one optimally responsive to *how things are* (as revealed in the evidence). If one can deliberately *alter* the evidence, this changes things completely. (Imagine a scientist who simply rejects any evidence not consistent with her favourite

¹⁹ I thank an anonymous reviewer for raising this objection.

theory). With prediction error minimization, however, *this is the whole point*: the set of priors that determine homeostasis are from the point of view of the organism's life *incontrovertible* in the face of countervailing evidence. As Bruineberg et al. (2016, 16) put it, “if my brain is a scientist, it is a crooked and fraudulent scientist” that is “heavily invested in ensuring the truth of a particular theory.” Worse, this means that *different* brains invest in the truth of *different theories*. Thus when Hohwy (2014) argues that prediction error minimization just *is* Bayesian inference, one must remember that it is at best Bayesian inference in a purely formal sense under conditions in which the overriding aim is to *change the evidence, not the priors*.

6 Predictive brains and social minds

If I am right, prediction error minimization cannot on its own explain the emergence of intentional states. As such, Hohwy's (2013, 1) contention that it explains “perception and action and everything mental in between” must be rejected. This has the following consequence: *if* one wants to hold onto predictive processing as our best theory of what the *brain* is doing, one must locate the source of intentionality *outside of the brain*. Of course, this is exactly the prescription we saw associated with pragmatism in Section 2.3: on this communitarian vision of mind, the semantic norms that ground intentional content arise through *normatively structured sociolinguistic practices*. As such, advocates of predictive processing *require* an account of intentional content of exactly the sort that pragmatism looks well-placed to provide.

Of course, I cannot fully deliver on this suggestion in the remaining space. Instead, I want to pick up from some remarks by Frith (2007) to provide some initial support and suggestions in its defence. Contemplating how the norm of *truth* might emerge in the context of predictive processing, Frith writes:

“In the very distant past our ancestors... were alone, constructing their models of the physical world, but unable to share them with others. At that time truth had no relevance for these models... All that mattered was that the model worked by predicting what would happen next” (Frith 2007, 179).

Here Frith identifies the basic lesson of Section 5: in the context of neuronal processes in the predictive brain, *truth* is irrelevant. He claims that what matters is *prediction*, but we saw in Section 4 that even this is not quite right: what really matters is homeostatic regulation (which in turn requires prediction). The basic point, however, is the same. How, then, do *our* internal states come to be answerable to this norm? Frith's (2007, 136) proposal is effectively pragmatist in spirit: “questions about the “truth” of the brain's models arise only when one brain *communicates* with another, and we discover that another person's model of the world is different from our own.”

In other words, the distinctive *truth* norm to which we hold *our* internal states responsible emerges in the idiosyncratic human context of social coordination made possible by *language*. Of course, this is exactly the communitarian vision of intentionality advocated in the pragmatist tradition since Dewey, and bears a striking resemblance to Price's (2011) proposal (see Section 2.3) that the truth norm arises as a “convenient friction” to facilitate *disagreement* and the wealth of socially adaptive

advantages that disagreement provides (cf. also Brandom 1994, ch.1). In the current context, we might put the idea as follows: natural language and shared cultural practices uniquely enable us to *externalise* our pragmatic predictive models and subject them to public critique, comparison and evaluation, a process that institutes a kind of socially constructed normativity against which we then hold our internal models and linguistic acts answerable (cf. Brandom 1994; Haugeland 1990).

Interestingly, Clark (2016) advances a similar suggestion in the last chapters of his recent monograph on the predictive mind. After outlining and exploring the core tenets and explanatory power of predictive processing, he raises a worry flagged by Roepstorff (2013) that the species-generalty of the processes it posits leaves it mysterious what explains the seemingly *novel* kinds of cognition exhibited by human beings. Clark's (2016) suggestion, following on from much work in the tradition of 4e cognition, is that what differentiates human cognition is *not* primarily what goes on within our heads, but rather the way in which *our* brains get augmented and transformed in the idiosyncratic environment of structured social interactions and public symbol systems characteristic of human life. As he puts it, "it is the predictive brain operating in rich bodily, social, and technological context that ushers minds like ours into the material realm." Of course, the explanation of our propensity to *inhabit* such social and symbolic environments must itself have a neural explanation. The suggestion, however, is "that [it was] some relatively *small* neural... difference [that] was the spark that lit a kind of intellectual forest fire" (Clark 2014, 179).

This suggestion exactly recapitulates Dewey's (1938/2008, 49) contention that the distinctive abilities exhibited in human intelligence are "instances of modifications wrought within the biological organism by the *cultural environment*." Nevertheless, Dewey and subsequent pragmatists go one step further: the proposal is not just that the distinctive sociocultural environment has ampliative and transformative effects on our cognitive *capacities*, but that it generates the very emergence of *intentional content* itself (Brandom 1994). It is the norms that emerge in such structured cultural contexts of evaluation and coordination to which the states of *our* internal models (i.e. pragmatic proxies)—uniquely in the terrestrial world—are answerable.

Of course, this proposal is extremely schematic, and much more work is required to show how previous proposals in the pragmatist tradition about the social construction of content could be integrated with predictive processing. Nevertheless, it suggests a surprising opportunity for mutual illumination between two frameworks that look for all the world to produce radically opposing visions of mentality. From predictive processing we learn the pragmatic origins of cognitive processes and the homeostatic construction of an affective niche, and from pragmatism we learn how predictive brains become transformed through structured processes of social coordination and evaluation—practices that produce the very norms from which the contents of our thoughts and utterances arise.

Although I have said little about what such an account might look like, I hope I have done enough to reveal it as a project *worth* pursuing. To reiterate, once one recognises the inadequacy of prediction error minimization to explain the distinctive *normativity* of intentionality, one must locate the source of that normativity elsewhere. I contend that pragmatism's communitarian vision of the human mind provides the "elsewhere": to understand how predictive brains acquire intentional states, one must situate them in the

broader environmental context of *human culture* and the qualitatively novel kinds of normativity it generates.

Before ending, I want to briefly flag two principled objections to this proposal—in the second case, as an impetus to the kinds of future work I think it provokes.

First, one might argue that sociocultural practices cannot play a *distinctive* explanatory role within the context of predictive processing, because the ability to participate in such practices must be explained by predictive processing itself. As such, the practices themselves are explanatorily redundant.²⁰ This objection confuses the conditions that enable predictive brains to *participate* in structured social practices for the practices themselves, however. Of course, no pragmatist has ever denied that something within the brain explains our capacity to participate in culture. The point is that such cultural practices produce phenomena that brains—whether predictive or otherwise—do not on their own.

A second and deeper objection is that the structured social practices I have appealed to cannot solve the problem I have asked of them. Specifically, the worry is that just as there is a mismatch between *prediction error minimization* and representational success, there is likewise a mismatch between social approval and truth (Boghossian 2006; Hutto and Satne 2015). As such, the proposal that *social norms* might ground content (e.g. Brandom 1994; Kripke 1982) falls foul of the very problem I have raised for predictive processing.

I cannot hope to address this enormous challenge to communitarian approaches to intentionality in the remaining space here, except to note that the pragmatist's response is always the same: *there is no alternative* (Rorty 1979, 1989). That is, the *world itself* cannot literally evaluate anything, and so socially constructed normativity is the best one can hope for when it comes to explaining the emergence of intentional content (Brandom 1994). For present purposes, the crucial point—stressed by Frith and many others—is that linguistically mediated social coordination between human beings gives rise to a genuinely *novel* and thus *distinct kind* of normativity from mere pragmatic success. It makes us answerable to *one* another—which, for a social species like us, effectively is the world (Rorty 1989). Elaborating this story—and exploring how the distinctive explanatory resources of predictive processing might alter and adapt that story in interesting ways—is a crucial project for future work.

7 Conclusion

I have argued that the initial appearance of a deep conflict between predictive processing and pragmatism as outlined in Sections 2 and 3 is not just illusory, but that something much more interesting is the case. Specifically, once one recognises the way in which predictive processing both vindicates and illuminates the first two pragmatist commitments outlined in Section 2, it becomes clear that the pragmatic, narcissistic character of prediction error minimization is inadequate to capture the distinctive normativity of intentionality. This effectively means that advocates of predictive processing *require* an extra-neural explanation of intentionality of just the sort that pragmatism's *communitarian* vision of human thought—the final and most

²⁰ I thank an anonymous reviewer for raising this worry.

controversial pragmatist commitment outlined in Section 2—can provide. Frith’s proposal about how the distinctive *truth* norm arises through communication provides some initial impetus in that direction, but fully substantiating it is of course a crucial task for future work. Nevertheless, I hope I have shown that it is work worth doing, and that the current paper has been a valuable contribution to the growing literature on predictive processing and its relationship to key debates in the philosophy of mind, cognitive science, and just about everything else.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Akins, K. (1996). Of sensory systems and the "Aboutness" of mental states. *The Journal of Philosophy*, 93(7), 337. <https://doi.org/10.2307/2941125>.
- Allen, M., & Friston, K. (2016). From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese*. <https://doi.org/10.1007/s11229-016-1288-5>.
- Anderson, M. (2017). Of Bayes and bullets: An embodied, situated, targeting-based account of predictive processing. In T. Metzinger & W. Wiese (Eds.), *philosophy and predictive processing*: 4. Frankfurt am main: MIND group. <https://doi.org/10.15502/9783958573055>.
- Anderson, M., & Chemero, T. (2013). The problem with brain GUTs: Conflation of different senses of "prediction" threatens metaphysical disaster. *Behavioral and Brain Sciences*, 36(03), 204–205. <https://doi.org/10.1017/s0140525x1200221x>.
- Bacon, M. (2012). *Pragmatism: An introduction* (1st ed.). Cambridge: Polity Press.
- Barrett, L. (2016). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, nsw154. <https://doi.org/10.1093/scan/nsw154>.
- Barrett, L. (2017). *How emotions are made* (1st ed.). New York: Macmillan.
- Bechtel, W. (2008). *Mental Mechanisms*. New York: Taylor & Francis Group.
- Bernstein, R. (2010). *The pragmatic turn* (1st ed.). Cambridge: Polity Press.
- Blackburn, S. (2006). *Truth: A guide for the perplexed* (1st ed.). London: Penguin.
- Boghossian, P. (2006). *Fear of knowledge*. Oxford: Clarendon.
- Brandom, R. (1994). *Making it explicit* (1st ed.). Cambridge: Harvard University Press.
- Brandom, R. (2002a). Pragmatics and pragmatisms. In J. Conant & U. Žegleń (Eds.), *Hilary Putnam: Pragmatism and realism* (1st ed., pp. 40–58). London: Routledge.
- Brandom, R. (2002b). *Tales of the Mighty Dead* (1st ed.). Cambridge, Mass. / London: Harvard University Press.
- Brandom, R. (2010). *Between saying and doing* (1st ed.). Oxford: Oxford University Press.
- Bruineberg, J., Kiverstein, J., & Rietveld, E. (2016). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese*. <https://doi.org/10.1007/s11229-016-1239-1>.
- Burge, T. (2010). *Origins of objectivity*. Oxford: OUP Oxford.
- Calvo, P., & Friston, K. (2017). Predicting green: Really radical (plant) predictive processing. *Journal of the Royal Society Interface*, 14(131), 20170096. <https://doi.org/10.1098/rsif.2017.0096>.
- Chemero, A. (2009). *Radical embodied cognitive science* (1st ed.). Cambridge: MIT Press.
- Clark, A. (2011). *Supersizing the mind*. New York: Oxford University Press.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03), 181–204. <https://doi.org/10.1017/s0140525x12000477>.
- Clark, A. (2014). *Mindware* (2nd ed.). New York: Oxford University Press.
- Clark, A. (2015). Predicting peace: The end of the representation wars—A reply to Michael Madary. In T. Metzinger & J. M. Windt (Eds.), *Open MIND*: 7(R). Frankfurt am Main: MIND Group. <https://doi.org/10.15502/9783958570979>.
- Clark, A. (2016). *Surfing uncertainty* (1st ed.). Oxford: Oxford University Press.

- CONANT, R., & ROSS ASHBY, W. (1970). Every good regulator of a system must be a model of that system †. *International Journal of Systems Science*, 1(2), 89–97. <https://doi.org/10.1080/00207727008920220>.
- Cummins, R. (1996). *Representations, targets, and attitudes*. Cambridge: The MIT Press.
- Denève, S., & Jardri, R. (2016). Circular inference: Mistaken belief, misplaced trust. *Current Opinion In Behavioral Sciences*, 11, 40–48. <https://doi.org/10.1016/j.cobeha.2016.04.001>.
- Dennett, D. (2013). *Intuition pumps and other tools for thinking* (1st ed.). London: Penguin Books.
- Dewey, J. (1925). *Experience and Nature* (1st ed.). La Salle: Open Court.
- Dewey, J. (1948). *Reconstruction in philosophy* (2nd ed.). Boston, Massachusetts: Beacon (Original Work Published 1920).
- Dewey, J. (2008). *Logic: The theory of inquiry*. In K. Poulous, *The Later Works* (1st ed.). Carbondale: Southern Illinois University Press.
- Egan, F. (2013). How to think about mental content. *Philosophical Studies*, 170(1), 115–135. <https://doi.org/10.1007/s11098-013-0172-0>.
- Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, 13(7), 293–301. <https://doi.org/10.1016/j.tics.2009.04.005>.
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>.
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, 29(1), 1–49. https://doi.org/10.1162/neco_a_00912.
- Frith, C. (2007). *Making Up The Mind: How the Brain Creates our Mental World* (1st ed.). Chichester: Wiley-Blackwell (an imprint of John Wiley & Sons Ltd).
- Gallagher, S., & Allen, M. (2016). Active inference, enactivism and the hermeneutics of social cognition. *Synthese*. <https://doi.org/10.1007/s11229-016-1269-8>.
- Gibson, J. (1979). *The ecological approach to visual perception* (1st ed.). Boston: Houghton-Mifflin.
- Gładziejewski, P. (2015). Predictive coding and representationalism. *Synthese*, 193(2), 559–582. <https://doi.org/10.1007/s11229-015-0762-9>.
- Godfrey-Smith, P. (2013). John Dewey's experience and nature. *Topoi*, 33(1), 285–291. <https://doi.org/10.1007/s11245-013-9214-7>.
- Godfrey-Smith, P. (2015). Pragmatism: Philosophical aspects. In J. Wright, *International Encyclopedia of the Social and Behavioral Sciences* (2nd ed., pp. 803–807). Oxford: Elsevier.
- Grice, H. (1957). Meaning. *The Philosophical Review*, 66(3), 377. <https://doi.org/10.2307/2182440>.
- Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27(03)
- Haugeland, J. (1990). The intentionality all-stars. *Philosophical Perspectives*, 4, 383–427. <https://doi.org/10.2307/2214199>.
- Heft, H. (2001). *Ecological psychology in context* (1st ed.). New Jersey: Psychology Press.
- Hobson, J., Hong, C., & Friston, K. (2014). Virtual reality and consciousness inference in dreaming. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01133>.
- Hohwy, J. (2013). *The predictive mind* (1st ed.). Oxford: Oxford University Press.
- Hohwy, J. (2014). The self-evidencing brain. *Nous*, 50(2), 259–285. <https://doi.org/10.1111/nous.12062>.
- Hohwy, J. (2015). The neural organ explains the mind. In T. Metzinger & J. M. Windt (Eds.), *OpenMIND*: 19(T). Frankfurt am Main: MIND Group. <https://doi.org/10.15502/9783958570016>.
- Hohwy, J. (2017). Priors in perception: Top-down modulation, Bayesian perceptual learning rate, and prediction error minimization. *Consciousness and Cognition*, 47, 75–85. <https://doi.org/10.1016/j.concog.2016.09.004>.
- Hutto, D. (2017). Getting into predictive processing's great guessing game: Bootstrap heaven or hell? *Synthese*. <https://doi.org/10.1007/s11229-017-1385-0>.
- Hutto, D., & Satne, G. (2015). The natural origins of content. *Philosophia*, 43(3), 521–536. <https://doi.org/10.1007/s11406-015-9644-0>.
- Hylton, P. (2007). *Quine* (1st ed.). New York: Routledge.
- James, W. (2000). *Pragmatism and other writings* (1st ed.). London: Penguin Books.
- Kiefer, A., & Hohwy, J. (2017). Content and misrepresentation in hierarchical generative models. *Synthese*. <https://doi.org/10.1007/s11229-017-1435-7>.
- Kirchhoff, M. (2017). Predictive processing, perceiving and imagining: Is to perceive to imagine, or something close to it? *Philosophical Studies*. <https://doi.org/10.1007/s11098-017-0891-8>.
- Kripke, S. (1982). *Wittgenstein on rules and private language*. Oxford: Oxford University Press.
- Lewis, D. (1975). Languages and language. In K. Gunderson (Ed.), *Minnesota studies in the philosophy of science* (1st ed., pp. 3–35). Minnesota: University of Minnesota Press.

- Mathys, C., Lomakina, E., Daunizeau, J., Iglesias, S., Brodersen, K., Friston, K., & Stephan, K. (2014). Uncertainty in perception and the hierarchical Gaussian filter. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00825>.
- McDowell, J. (1994). *Mind and world*. Cambridge: Harvard University Press.
- Menary, R. (2015). Pragmatism and the pragmatic turn in cognitive science. In A. Engel, K. Friston, & D. Kragic (Eds.), *The pragmatic turn: Toward action-oriented views in cognitive science* (1st ed., pp. 215–235). London: MIT Press.
- Millikan, R. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
- Peirce, C., Hartshorne, C., & Weiss, P. (1974). *Collected papers of Charles Sanders Peirce* (1st ed.). Cambridge: Belknap Press.
- Pendl, S., Salzwedel, A., Goldman, B., Barrett, L., Lin, W., Gilmore, J., & Gao, W. (2017). Emergence of a hierarchical brain during infancy reflected by stepwise functional connectivity. *Human Brain Mapping*, 38(5), 2666–2682. <https://doi.org/10.1002/hbm.23552>.
- Penny, W. (2012). Bayesian models of brain and behaviour. *ISRN Biomathematics*, 2012, 1–19. <https://doi.org/10.5402/2012/785791>.
- Price, H. (2011). *Naturalism without mirrors* (1st ed.). New York: Oxford University Press.
- Price, H. (2013). *Expressivism, Pragmatism and Representationalism* (1st ed.). Cambridge University Press.
- Quine, W. (1960). *Word and object* (1st ed.). Cambridge: MIT Press.
- Ramsey, W. (2007). *Representation reconsidered*. Cambridge: Cambridge University Press.
- Rescorla, M. (2013). Bayesian perceptual psychology. In M. Matthen (Ed.), *Oxford handbook of the philosophy of perception*. Oxford: Oxford University Press.
- Rockwell, T. (2005). *Neither brain nor ghost* (1st ed.). Cambridge: MIT Press.
- Roepstorff, A. (2013). Interactively human: Sharing time, constructing materiality. *Behavioral and Brain Sciences*, 36(03), 224–225. <https://doi.org/10.1017/s0140525x12002427>.
- Rorty, R. (1979). *Philosophy and the mirror of nature* (1st ed.). Princeton: Princeton University Press.
- Rorty, R. (1989). *Contingency, irony, and solidarity* (1st ed.). Cambridge [u.a.]: Cambridge Univ. Press.
- Rorty, R. (1999). *Philosophy and social hope* (1st ed.). London: Penguin.
- Satne, G. (2017). Social approaches to intentionality. In J. Kiverstein (Ed.), *The Routledge Handbook of Philosophy of the Social Mind* (1st ed., pp. 528–544). London: Routledge.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(03), 417. <https://doi.org/10.1017/s0140525x00005756>.
- Sellars, W. (1963). *Science, perception and reality* (1st ed.). New York: The Humanities Press.
- Seth, A. K. (2015). The cybernetic bayesian brain—From interoceptive inference to sensorimotor contingencies. In T. Metzinger & J. M. Windt (Eds.), *Open MIND: 35(T)*. Frankfurt: MIND Group. <https://doi.org/10.15502/9783958570108>.
- Sims, A. (2016). A problem of scope for the free energy principle as a theory of cognition. *Philosophical Psychology*, 29(7), 967–980. <https://doi.org/10.1080/09515089.2016.1200024>.
- Sterling, P., & Laughlin, S. (2015). *Principles of neural design* (1st ed.). London: MIT Press.
- Tenenbaum, J., Kemp, C., Griffiths, T., & Goodman, N. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022), 1279–1285. <https://doi.org/10.1126/science.1192788>.
- Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind* (1st ed.). London: Harvard University Press.
- Varela, F., Thompson, E., Rosch, E., & Kabat-Zinn, J. (1991). *The embodied mind* (1st ed.). Cumberland: MIT Press.
- Williams, D. (2017). Predictive processing and the representation wars. *Minds And Machines*. <https://doi.org/10.1007/s11023-017-9441-6>.
- Williams, D., & Colling, L. (2017). From symbols to icons: The return of resemblance in the cognitive neuroscience revolution. *Synthese*. <https://doi.org/10.1007/s11229-017-1578-6>.